

UPPER BOUND FOR INTERMEDIATE SINGULAR VALUES OF RANDOM MATRICES

FENG WEI

ABSTRACT. In this paper, we prove that an $n \times n$ matrix A with independent centered subgaussian entries satisfies

$$s_{n+1-l}(A) \leq C_1 t \frac{l}{\sqrt{n}}$$

with probability at least $1 - \exp(-C_2 tl)$. This yields $s_{n+1-l}(A) \sim \frac{cl}{\sqrt{n}}$ in combination with a known lower bound. These results can be generalized to the rectangular matrix case.

1. INTRODUCTION

Consider an $n \times m$ real matrix A with $n \geq m$. The singular values $s_k(A)$ of A , where $k = 1, 2, \dots, n$, are the eigenvalues of $\sqrt{A^T A}$ arranged in non-increasing order. The non-asymptotic singular value distribution of random i.i.d. sub-gaussian matrix is an important and interesting subfield in random matrix theory. The first result in this direction was obtained in [4], where it was proved that the smallest singular value of a square i.i.d. sub-gaussian matrix is bounded below by $n^{-3/2}$ with high probability. This result was later extended and improved in a number of papers, including [18, 19, 10, 1, 6]. The above mentioned results pertain to square matrices. A probabilistic lower bound for the smallest singular value of a rectangular matrix was obtained by M. Rudelson and R. Vershynin [11]. They proved that an $n \times (n-l)$ matrix has smallest singular value lower bounded by $\frac{\varepsilon l}{\sqrt{n}}$ with probability at least $1 - (C\varepsilon)^l - \exp(-Cn)$. Using this result, one can show that for a square i.i.d. sub-gaussian matrix A , $s_{n+1-l}(A) > c \frac{l}{\sqrt{n}}$ with high probability.

However, the optimal upper bound of the singular values for general sub-gaussian matrices is unknown. Prior to this paper, the only progress in this direction was made by Szarek [17] who proposed an optimal upper bound for the gaussian matrix. Szarek proved that for

Partially supported by M. Rudelson's NSF Grant DMS-1464514, and USAF Grant FA9550-14-1-0009.

a standard gaussian i.i.d. matrix G , $\frac{Cl}{\sqrt{n}} \leq s_{n+1-l}(G) \leq \frac{Cl}{\sqrt{n}}$ with probability at least $1 - \exp(-Cl^2)$. This result suggests that l th smallest singular value of an i.i.d. sub-gaussian matrix is concentrated around $\frac{l}{\sqrt{n}}$.

Although the optimal upper bound is not proved for general matrices, some results can be deduced. T. Tao and V. Vu have established the universal behavior of small singular values in [19] (see Theorem 6.2 [19]). Combined with Szarek's Theorem 1.3 in [17], their approach allows us to deduce some non-asymptotic bounds for random i.i.d. square matrix under a moment condition. However, their bound only works for $l \leq n^c$ where c is a small constant. Tao and Vu's approach [19] is based on the Berry-Esseen Theorem for the frames and does not allow one to obtain exponential bounds for the probability as we do in our Theorem 1.6. Also, C. Cacciapuoti, A. Maltsev, B. Schlein estimated the rate of convergence of the empirical measure of singular values to the limit distribution near the hard edge (see [2] Theorem 3). Theorem 3 in [2] can be used to derive an upper bound of the form $\frac{cl^C}{\sqrt{n}}$ [2]. Better understood is the upper bound for the smallest singular value. M. Rudelson and R. Vershynin were the first to prove the smallest singular value of the i.i.d. sub-gaussian matrix is also bounded from above by $\frac{c}{\sqrt{n}}$ with high probability (see [9]). A different proof with an exponential tail probability can be found in a very recent paper by H. Nguyen and V. Vu [5].

In this paper, we prove the upper bound on the singular values under two assumptions: that the entries of the matrix are non-degenerate; and that they have a fast tail decay. The first assumption is quantified in terms of the Levy concentration function and the second is quantified in terms of the ψ_θ -norm. Next we provide definitions.

Definition 1.1. Let Z be a random vector that takes values in \mathbb{C}^n . The concentration function of Z is defined as

$$\mathcal{L}(Z, t) = \sup_{u \in \mathbb{C}^n} \mathbb{P}\{\|Z - u\|_2 \leq t\}, t \geq 0.$$

Definition 1.2. Let $\theta > 0$. Let Z be a random variable on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Then the ψ_θ -norm of Z is defined as

$$\|Z\|_{\psi_\theta} := \inf \left\{ \lambda > 0 : \mathbb{E} \exp \left(\left(\frac{|Z|}{\lambda} \right)^\theta \right) \leq 2 \right\}$$

If $\|X\|_{\psi_\theta} < \infty$, then X is called a ψ_θ random variable. This condition is satisfied for broad classes of random variables. In particular, a

bounded random variable is ψ_θ for any $\theta > 0$, a normal random variable is ψ_2 , and a Poisson variable is ψ_1 .

In this paper, we prove that for all l , $s_{n+1-l}(A) \geq \frac{Ct}{\sqrt{n}}$ with an exponentially small probability, where A is a random matrix under the following assumption:

Assumption 1.3. *Let $p > 0$. Let A be an $n \times m$ random matrix with i.i.d. entries that have mean 0, variance 1 and ψ_2 -norm K . Assume also that there exists $0 < s \leq s_0(p, K)$ such that*

$$\mathcal{L}(A_{i,j}, s) \leq ps.$$

Here, $s_0(p, K)$ is a given function depending only on p and K .

A concrete value of $s_0(p, K)$ can be determined by tracing the proof of Theorem 1.6.

Remark 1.4. The condition on the Levy concentration function is automatically satisfied if the density of the entries is bounded by p . However, our result holds in a much more general setting because we require this condition to hold only for one fixed s and not for all $s > 0$. This assumption can be viewed as a discrete analog of the bounded density condition.

Remark 1.5. The analysis of the proof for Theorem 1.6 shows that it is enough to take $s_0(p, K) = c(K) \min\{p^{-1}, 1\}$, where $c(K)$ is a small constant that depends only on K .

We prove the following main theorem for a random matrix A satisfying Assumption 1.3:

Theorem 1.6. (Upper bound for singular values of an i.i.d. sub-gaussian square matrix) *Let A be an $n \times n$ random matrix that satisfies Assumption 1.3 with some $s_0(p, K)$ that depends only on p, K . Then there exist constants $C_1, C_2 > 0$ such that for all $t > 1$, for all l between 1 and n ,*

$$\mathbb{P} \left(s_{n+1-l}(A) \leq C_1 \frac{t}{\sqrt{n}} \right) \geq 1 - \exp(-C_2 t)$$

where C_1, C_2 are constants that depend only on K, p .

Remark 1.7. In [5], Nguyen and Vu obtained a sharp bound for the smallest singular value of i.i.d. sub-gaussian matrices with exponential tail probability. Theorem 1.6 recovers the result obtained by Nguyen and Vu under Assumption 1.3 and generalizes that result to all l .

Remark 1.8. Unlike the probability bound deduced from [19], our probability tail bound is exponential type. Szarek's probability estimate [17] suggests that the optimal probability bound for $\mathbb{P}\left(s_{n+1-l}(A) \leq \frac{cl}{\sqrt{n}}\right)$ is $1 - \exp(-Cl^2)$.

Remark 1.9. Possible applications of Theorem 1.6 include eigenvector l_∞ delocalization of random matrices [13]. For that, one has to consider $A - zI$ instead of the matrix A . Further effort would be needed to generalize our result to the case with a shift.

Also, Theorem 1.6 can be extended to rectangular matrices easily (see Section 4). And more precisely, we have the following corollary:

Corollary 1.10. (Upper bound for singular values of an i.i.d. sub-gaussian matrix) *Let A be an $n \times (n - k)$ random matrix that satisfies Assumption 1.3 with some $s_0(p, K)$ that depends only on p, K . Then there exist constants $C_1, C_2 > 0$ such that for all $t > 1$ and l between 1 and n ,*

$$\mathbb{P}\left(s_{n+1-l}(A) \leq C_1 \frac{tl}{\sqrt{n}}\right) \geq 1 - \exp(-C_2 tl)$$

where C_1, C_2 are constants that depend only on K, p .

A direct application of Theorem 1.6 and Theorem 2.6 leads us to a generalization of Theorem 1.3 in [17].

Corollary 1.11. (Non-asymptotic singular values distribution of i.i.d. sub-gaussian square matrix) *Let A be an $n \times n$ random matrix that satisfies Assumption 1.3 with some $s_0(p, K)$ that depends only on p, K . Then there exist $0 < C_1 < C_2$ and $C_3 > 0$, such that for all l between 1 and n ,*

$$\mathbb{P}\left(\frac{C_1 l}{\sqrt{n}} \leq s_{n+1-l}(A) \leq \frac{C_2 l}{\sqrt{n}}\right) \geq 1 - \exp(-C_3 l)$$

where C_i s are constants that depend only on K, p .

A similar proof can lead to an analog in the rectangular case:

Corollary 1.12. (Non-asymptotic distribution of singular values in the i.i.d. sub-gaussian rectangular matrix) *Let A be an $n \times (n - k)$ random matrix that satisfies Assumption 1.3 with some $s_0(p, K)$ that depends only on p, K . Then there exist $0 < C_1 < C_2$ and $C_3 > 0$, such that for all l between k and n ,*

$$\mathbb{P}\left(\frac{C_1 l}{\sqrt{n}} \leq s_{n+1-l}(A) \leq \frac{C_2 l}{\sqrt{n}}\right) \geq 1 - \exp(-C_3 l)$$

where C_i s are constants that depend only on K, p .

In Section 2, we present some preliminary results that are needed to prove Theorem 1.6. In Section 3, we provide proof of Theorem 1.6. We prove Theorems 1.10 and 1.11 In Section 4.

2. NOTATION AND PRELIMINARIES

Throughout this paper, c denotes absolute constants, C denotes constants that may depend only on the parameters K, p . Note that these constants may vary from line to line. S^{n-1} denotes the n dimensional sphere, i.e., the sphere in \mathbb{R}^n which itself is an $(n-1)$ -dimensional manifold. S_E denotes the sphere of a subspace E , i.e., $S_E = S^{n-1} \cap E$.

For a $n \times n$ random matrix A , as in Theorem 1.6, we denote by A_l the $n \times l$ matrices of the first l columns of A . A_{n-l} denotes the $n \times (n-l)$ matrix of the last $n-l$ columns of A . Without loss of generality, we can assume A is a.s. invertible. If not we prove the theorem for $A + \varepsilon G$, where G is an independent gaussian matrix. Then the result holds for A up to an absolute constant by sending ε to zero. X_k will denote the k th column of matrix A , and we use the following notations

- $H_l := \text{span}(X_j)_{j>l}$.
- $H_{l,k} := \text{span}(X_j)_{j>l, j \neq k}$.
- $P_l, P_{l,k}$ are the orthogonal projections onto $H_l, H_{l,k}$, respectively.
- $P_l^\perp, P_{l,k}^\perp$ are the orthogonal projections onto $H_l^\perp, H_{l,k}^\perp$, respectively.
- $X_k^* := (A^{-1})^* e_k$, i.e., the k -th column of $(A^{-1})^*$.
- $Y_k^* := P_l X_k^*, k = l+1, l+2, \dots, n$.

2.1. Biorthogonal system. Consider vectors $(v_k)_{k=1}^n$ and $(v_k^*)_{k=1}^n$ that belong to an n -dimensional Hilbert space H . Recall that the system $(v_k, v_k^*)_{k=1}^n$ is called a biorthogonal system in H if $\langle v_j, v_k^* \rangle = \delta_{j,k}$ for all $j, k = 1, 2, \dots, n$. The system is called complete if $\text{span}(v_k) = H$. The following proposition summarizes some elementary known properties of biorthogonal systems. For the reader's convenience, we provide the proof of this elementary fact in the appendix.

Proposition 2.1. (1) *Let D be an $n \times n$ invertible matrix with columns $v_k = D e_k, k = 1, 2, \dots, n$. Define $v_k^* = (D^{-1})^* e_k$. Then $(v_k, v_k^*)_{k=1}^n$ is a complete biorthogonal system in \mathbb{R}^n .*

- (2) *Let $(v_k)_{k=1}^n$ be a linearly independent system in an n -dimensional Hilbert space H . Then there exist unique vectors $(v_k^*)_{k=1}^n$ such that $(v_k, v_k^*)_{k=1}^n$ is a biorthogonal system in H . This system is complete.*

- (3) Let $(v_k, v_k^*)_{k=1}^n$ be a complete biorthogonal system in a Hilbert space H . Then $\|v_k^*\|_2 = 1/\text{dist}(v_k, \text{span}(v_j)_{j \neq k})$ for $k = 1, 2, \dots, n$.

The next Lemma tells us the relation between Y_k^* and X_k^* for $k \geq l+1$.

Lemma 2.2. $(X_k, Y_k^*)_{k=l+1}^n$ is a complete biorthogonal system in H_l .

Proof. By definition, for all $k \geq l+1$

$$Y_k^* - X_k^* \in \ker(P_l) = H_l^\perp = \text{span}(X_j^*)_{j \leq l}.$$

So we have, for all $k \geq l+1$

$$Y_k^* = X_k^* + \sum_{j=1}^l a_{kj} X_j^*, \text{ for some } a_{kj} \in \mathbb{R}, j = 1, 2, \dots, l.$$

By the orthogonality, we have for all $k, i \geq l+1$

$$\langle Y_k^*, X_i \rangle = \langle X_k^*, X_i \rangle + \sum_{j=1}^l a_{kj} \langle X_j^*, X_i \rangle = \langle X_k^*, X_i \rangle = \delta_{k,i}.$$

Thus the biorthogonality is proved. The completeness follows since $\dim(H_l) = n - l$. \square

In view of the uniqueness of Part 2 of Proposition 2.1, Lemma 2.2 has the following crucial consequence.

Corollary 2.3. The system of vectors $(Y_k^*)_{k=l+1}^n$ is uniquely determined by the system $(X_k)_{k=l+1}^n$. In particular, the random vector system $(Y_k^*)_{k=l+1}^n$ is independent with random vector system $(X_k)_{k=1}^l$.

2.2. Concentration theorems. The major tools of our proof come from measure concentration theory. Here we list the concentration theorems that will be used in the proof.

The first theorem is a concentration property of sub-gaussian random vectors.

Theorem 2.4. Let D be a fixed $m \times n$ matrix. Consider a random vector Z with independent entries that have mean 0, variance greater than 1, and uniformly bounded by K in ψ_2 norm. Then, for any $t \geq 0$, we have

$$\mathbb{P}(|\|DZ\|_2 - M| > t) \leq 2 \exp\left(-\frac{ct^2}{\|D\|^2}\right)$$

where $M = (\mathbb{E}\|DZ\|_2^2)^{1/2}$ which satisfies $\|D\|_{\text{HS}} \leq M \leq K\|D\|_{\text{HS}}$, and $c = c(K)$ is polynomial in K .

This result can be deduced from the Hanson-Wright inequality. A modern proof of the Hanson-Wright inequality and a deduction of the above Theorem 2.4 are discussed in [14].

Sub-gaussian concentration paired with a standard covering argument yields the following result on norms of random matrices, see [14].

Theorem 2.5. (Products of random and deterministic matrices). *Let D be a fixed $m \times N$ matrix, and let G be an $N \times k$ random matrix with independent entries that satisfy $\mathbb{E}G_{ij} = 0$, $\mathbb{E}G_{ij}^2 \geq 1$ and $\|G_{ij}\|_{\psi_2} \leq K$. Then for any $s, t \geq 1$ we have*

$$\mathbb{P}\{\|DG\| > C(s\|D\|_{\text{HS}} + t\sqrt{k}\|D\|)\} \leq 2\exp(-s^2r - t^2k)$$

Here $r = \|D\|_{\text{HS}}^2/\|D\|_2^2$ is the stable rank of D , and $C = C(K)$ is a polynomial in K .

The following result gives the lower bound on the smallest singular value of a rectangular i.i.d. sub-gaussian matrix. This will be used in our proof of Theorem 1.6; it can also yield the lower bound in Theorem 1.11 directly. The proof and extensions of the theorem are discussed in [4, 10, 11, 12].

Theorem 2.6. *Let G be an $N \times n$ random matrix, $N \geq n$, whose elements are independent copies of a centered sub-gaussian random variable with unit variance. Then for every $\varepsilon > 0$, we have*

$$\mathbb{P}\left(s_n(G) \leq \varepsilon \left(\sqrt{N} - \sqrt{n-1}\right)\right) \leq (C\varepsilon)^{N-n+1} + e^{-C'N}$$

where $C, C' > 0$ depend (polynomially) only on the sub-gaussian moment K .

As one step towards the above least singular value bound, the following distance to a random subspace theorem was proved by M. Rudelson and R. Vershynin [11]:

Theorem 2.7. (Distance to a random subspace). *Let Z be a vector in \mathbb{R}^N whose coordinates are independent and identically distributed centered sub-gaussian random variables with unit variance. Let H be a random subspace in \mathbb{R}^N spanned by $N - m$ vectors, $0 < m < \tilde{c}N$, whose coordinates are independent and identically distributed centered sub-gaussian random variables with unit variance, independent of Z . Then, for every $v \in \mathbb{R}^N$ and every $\varepsilon > 0$, we have*

$$\mathbb{P}(\text{dist}(Z, H + v) < \varepsilon\sqrt{m}) \leq (C\varepsilon)^m + e^{-cN},$$

where C, c, \tilde{c} depend only on the sub-gaussian moments.

M. Rudelson and R. Vershynin have recently proved the following results for small ball probability of a linear image of high dimensional distributions [15] (see also [3]).

Theorem 2.8. (Concentration function of projections.) *Consider a random vector $Z = (Z_1, \dots, Z_n)$ where Z_i are real-valued independent random variables. Let $t, p \geq 0$ be such that*

$$\mathcal{L}(Z_i, t) \leq p \text{ for all } i = 1, \dots, n$$

Let P be an orthogonal projection in \mathbb{R}^n onto a d -dimensional subspace. Then

$$\mathcal{L}(PZ, t\sqrt{d}) \leq (cp)^d.$$

where c is an absolute constant.

In the same paper, Rudelson and Vershynin generalized Theorem 2.8 to general matrices:

Theorem 2.9. (Concentration functions of anisotropic distributions.) *Consider a random vector Z where Z_i are real-valued independent random variables. Let $t, p \geq 0$ be such that*

$$\mathcal{L}(Z_i, t) \leq p \text{ for all } i = 1, \dots, n$$

Let D be an $m \times n$ matrix and $\varepsilon \in (0, 1)$. Then

$$\mathcal{L}(DZ, t\|D\|_{\text{HS}}) \leq (c_\varepsilon p)^{(1-\varepsilon)r(D)}$$

where $r(D) = \|D\|_{\text{HS}}^2 / \|D\|_2^2$ and c_ε depend only on ε .

As a special case of Theorem 2.8, the following corollary usefully controls the concentration function of sums:

Theorem 2.10. (Concentration function of sums.) *Consider a random vector $Z = (Z_1, \dots, Z_n)$ where Z_i are real-valued independent random variables. Let $t, p \geq 0$ be such that*

$$\mathcal{L}(Z_i, t) \leq p \text{ for all } i = 1, \dots, n$$

Let a_1, \dots, a_n be real numbers with $\sum_{j=1}^n a_j^2 = 1$. Then

$$\mathcal{L}\left(\sum_{i=1}^n a_i Z_i, t\right) \leq cp.$$

where c is an absolute constant.

3. PROOF OF THEOREM 1.6

Before proving the theorem, let us explain our strategy. We prove a lower bound for $s_l(A^{-1})$, rather than proving an upper bound for $s_{n+1-l}(A)$. To do this, we show that there exists an l dimensional subspace, such that the smallest singular value of the operator restricted on this subspace is bounded from below. Our target subspace will be H_l^\perp .

The proof uses a net argument for a specially constructed net. In Step 1, we obtain a small ball probability estimate for a random vector. A generic vector in H_l^\perp can be represented as $A^{-1}P_l^\perp A_l y$ for some $y \in \mathbb{R}^l$. We will show that $\frac{\|A^{-1}P_l^\perp A_l y\|_2}{\|P_l^\perp A_l y\|_2}$ is bounded from below by $\frac{C\sqrt{n}}{l}$ for any $y \in S^{l-1}$. In steps 2 and 3, we provide a union bound argument.

There are three essential features of our proof. First, as H_l^\perp is a random subspace, we cannot consider a net on $S_{H_l^\perp}$ directly. So, we consider a net \mathcal{N}_ε on S^{l-1} instead, which will induce a net on H_l^\perp . Second, to complete the argument we need to show the union bound probability of the form $|\mathcal{N}_\varepsilon| \exp(-Cl)$ is small, where C is a small constant. Since $|\mathcal{N}_\varepsilon| \sim \left(\frac{3}{\varepsilon}\right)^l$, this bound in general can be large. To control the probability, we work not on $y \in S^{l-1}$ but on $y \in S^{l'-1}$, $l' = \lfloor \kappa l \rfloor$ for some $\kappa \in (0, 1)$ instead. With this dimension reduction argument, we end up proving that $s_{l'}(A^{-1}) \geq \frac{C\sqrt{n}}{l}$, and then we rephrase it. Third, representing a vector from H_l^\perp as $A^{-1}P_l^\perp A_l y$ is advantageous because

$$\|A^{-1}P_l^\perp A_l y\|_2^2 = \|B A_l y\|_2^2 + 1$$

where B is a random matrix that is independent of A_l . This allows us to analyze the property of B first and then apply tools like Theorem 2.9 and Theorem 2.5. Note that this construction was generalized from the one dimensional case presented by M. Rudelson and R. Vershynin [9].

In the proof, we will use the well-known estimate that there exists an ε -net on S^{l-1} with cardinality less than $\left(\frac{3}{\varepsilon}\right)^l$, see, e.g., Lemma 4.3 in [7].

Proof of Theorem 1.6. To prove Theorem 1.6, we only need to prove the following claim:

Claim. There exist C_1 and C_2 that only depend on K, p such that for every l between 1 and n ,

$$\mathbb{P} \left(s_{n+1-l}(A) \geq C_1 \frac{l}{\sqrt{n}} \right) \leq \exp(-C_2 l).$$

To start, we derive Theorem 1.6 from the claim. Let $t \geq 1$, and let k be any integer between 1 and n . Set $l = \lfloor tk \rfloor$ and assume for a moment that $l < n$. Then

$$(1) \quad \mathbb{P} \left(s_{n+1-k}(A) \geq C_1 \frac{2tk}{\sqrt{n}} \right) \leq \mathbb{P} \left(s_{n+1-k}(A) \geq \frac{C_1 l}{\sqrt{n}} \right) \leq \exp(-C_2 l) \leq \exp \left(-\frac{C_2 tk}{2} \right).$$

In the case $l \geq n$, the sub-gaussian tail estimate for the norm of a random matrix (one may also consider this as a special case of Theorem 2.5) yields

$$(2) \quad \mathbb{P} \left(s_{n+1-k}(A) \geq C_3 \frac{2tk}{\sqrt{n}} \right) \leq \mathbb{P} \left(s_1(A) \geq C_3 \frac{2tk}{\sqrt{n}} \right) \leq \exp \left(-C \frac{C_3^2 t^2 k^2}{n} \right) \leq \exp(-C_4 tk),$$

and therefore for all k between 1 and n ,

$$\mathbb{P} \left(s_{n+1-k}(A) \geq C_5 \frac{tk}{\sqrt{n}} \right) \leq \exp(-C_6 tk)$$

with constants C_5, C_6 depending on p, K only. So Theorem 1.6 is implied by the claim. \square

Now, we prove the above claim.

Proof of the claim. In the proof of the claim, we first assume $l \leq \frac{\tilde{c}n}{2}$, where \tilde{c} is the same as the \tilde{c} which appeared in Theorem 2.7. If $l > \frac{\tilde{c}n}{2}$, then the required bound follows from the estimate for $s_1(A)$. This is a standard estimate of the operator norm that can be found in many places, for example, in Theorem 2.4 of [11]. Let $\alpha > 1, \delta, \kappa < 1, \beta < \alpha^{-1} < 1$ be parameters to be chosen later. Also, assume that

$$(3) \quad \mathcal{L}(A_{i,j}, \beta) \leq p\beta$$

i.e. Assumption 1.3 is true with $s = s_0(p, K) = \beta$.

Step 1. Concentration for a random vector. Consider $y \in S^{l-1}$, define

$$(4) \quad U(y) := X(y) - P_l X(y) := A_l y - P_l A_l y.$$

then $X(y) := A_l y$ is still a mean 0, variance 1 sub-gaussian random vector. According to the Hoeffding inequality, the sub-gaussian moment of entries of $X(y)$ is bounded above by CK (see Theorem 3.3 in [7]). Without ambiguity, we use the notation U, X instead of $U(y), X(y)$.

In step 1, we show with high probability that

$$\|U\|_2 \lesssim \sqrt{l}, \quad \|A^{-1}U\|_2 \gtrsim \frac{\sqrt{n}}{\sqrt{l}}.$$

First, we give an upper bound for $\|P_l^\perp A_l\|$. This leads to an uniform upper bound of $\|U(y)\|_2$ for all $y \in S^{l-1}$.

Step 1.1. Concentration of $\|P_l^\perp A_l\|$.

First, notice that $I - P_l = P_l^\perp$, which is an orthogonal projection onto H_l^\perp , so it does not depend on A_l only on A_{n-l} . Thus, P_l^\perp can be treated as a fixed matrix. We apply Theorem 2.5 with $B = P_l^\perp$ and $G = A_l$, then we have

$$\mathbb{P}(\|P_l^\perp A_l\| > \alpha\sqrt{l}) \leq 2\exp(-C\alpha^2 l).$$

In particular, for a single vector we have

$$\mathbb{P}(\|U\|_2 > \alpha\sqrt{l}) = \mathbb{P}(\|P_l^\perp A_l y\|_2 > \alpha\sqrt{l}) \leq 2\exp(-C\alpha^2 l).$$

Step 1.2. Concentration of $\|A^{-1}U\|_2$.

Now consider

$$A^{-1}U = A^{-1}A_l y - A^{-1}P_l A_l y = y - A^{-1}P_l A_l y.$$

Notice that $A^{-1}P_l A_l y$ is supported in $\text{span}\{e_{l+1}, \dots, e_n\}$ because $P_l A_l y \in H_l$. So we have

$$\begin{aligned} \|A^{-1}U\|_2^2 &= \|y\|_2^2 + \|A^{-1}P_l A_l y\|_2^2 > \|A^{-1}P_l A_l y\|_2^2 \\ &= \sum_{k=1}^n \langle A^{-1}P_l A_l y, e_k \rangle^2 = \sum_{k=1}^n \langle P_l X, (A^{-1})^T e_k \rangle^2 \\ (5) \quad &= \sum_{k=1}^n \langle P_l X, X_k^* \rangle^2 = \sum_{k=l+1}^n \langle P_l X, X_k^* \rangle^2 \\ &= \sum_{k=l+1}^n \langle X, P_l X_k^* \rangle^2 = \sum_{k=l+1}^n \langle X, Y_k^* \rangle^2 \end{aligned}$$

where in the third line we used the fact that $(X_k, X_k^*)_{k=1}^n$ forms a complete biorthogonal system on \mathbb{R}^n from Lemma 2.1. Thus, $X_k^* \perp H_l, k \leq l$.

Using the above property, let B be the $(n-l) \times n$ matrix whose rows are $(Y_k^*)^T, k = l+1, l+2, \dots, n$. Then we have

$$\|A^{-1}U\|_2^2 \geq \sum_{k=l+1}^n \langle X, Y_k^* \rangle^2 = \|BX\|_2^2.$$

Our goal is to get a small ball probability estimate of $\|BX\|$. As B is independent of X , we would like to apply Theorem 2.9. Thus, we first need an estimate for $\|B\|$ and $\|B\|_{\text{HS}}$.

Step 1.2.1. Lower bound of $\|B\|_{\text{HS}}$.

According to Theorems 2.3, 2.2 and 2.1, we have

$$\|B\|_{\text{HS}}^2 = \sum_{k=l+1}^n \|Y_k^*\|_2^2 = \sum_{k=l+1}^n \text{dist}(X_k, H_{l,k})^{-2} = \sum_{k=l+1}^n \|P_{l,k}^\perp X_k\|_2^{-2}.$$

Denote $V_j = \text{dist}^2(X_j, H_{l,j})$. Then

$$\begin{aligned} (6) \quad \mathbb{P} \left(\|B\|_{\text{HS}} < \alpha^{-1} \sqrt{\frac{n-l}{l}} \right) &= \mathbb{P} \left(\left(\frac{1}{n-l} \sum_{j=l+1}^n V_j^{-1} \right)^{-1} > \alpha^2 l \right) \\ &\leq \mathbb{P} \left(\frac{1}{n-l} \sum_{j=l+1}^n V_j > \alpha^2 l \right) \\ &= \mathbb{P} \left(\frac{1}{n-l} \sum_{j=l+1}^n (V_j - 4(l+1)) > \alpha^2 l - 4(l+1) \right) \\ &\leq \mathbb{P} \left(\frac{1}{n-l} \sum_{j=l+1}^n (V_j - 4(l+1))_+ > \frac{\alpha^2}{2} l \right). \end{aligned}$$

where the first inequality follows from the inequality between harmonic mean and arithmetic mean and the second inequality is trivial if we provide $\alpha^2 > 10$. Consider

$$\begin{aligned} (7) \quad \mathbb{P}((V_j - 4(l+1))_+ > 4t) &\leq \mathbb{P}(\sqrt{V_j} > \sqrt{l+1} + \sqrt{t}) \\ &= \mathbb{P}(\|P_{l,j}^\perp X_j\|_2 - \sqrt{l+1} > \sqrt{t}) \end{aligned}$$

Then applying Theorem 2.4 with $A = P_{l,k}^\perp$ we have $M = \sqrt{l+1}$. Thus $(V_j - 4(l+1))_+$ is a sub-exponential random variable with $\|(V_j - 4(l+1))_+\|_{\psi_1} \leq C$. By the triangle inequality,

$$\left\| \frac{1}{n-l} \sum_{j=l+1}^n (V_j - 4(l+1))_+ \right\|_{\psi_1} \leq C.$$

Recalling that $l \leq \frac{n}{2}$, we have

$$\mathbb{P} \left(\|B\|_{\text{HS}} < \alpha^{-1} \sqrt{\frac{n}{l}} \right) \leq \exp(-C\alpha^2 l).$$

Step 1.2.2. Upper bound of $\|B\|$.

First, we have

$$\begin{aligned}
 \|B\|^2 &= \sup_{x \in S^{n-1} - \{0\}} \|Bx\|_2^2 = \sup_{x \in \mathbb{R}^n - \{0\}} \frac{\|Bx\|_2^2}{\|x\|_2^2} = \sup_{\|Bx\|_2=1} \frac{1}{\|x\|_2^2} \\
 (8) \quad &= \sup_{\substack{x \in \mathbb{R}^n - \{0\} \\ \sum_{k=l+1}^n \langle x, Y_k^* \rangle^2 = 1}} \frac{1}{\|x\|_2^2} = \sup_{\substack{x \in H_l - \{0\} \\ \sum_{k=l+1}^n \langle x, Y_k^* \rangle^2 = 1}} \frac{1}{\|x\|_2^2}
 \end{aligned}$$

where the last equality can be justified by considering the decomposition $x = x' + x''$, $x' \in H_l$, $x'' \in H_l^\perp$ with $\|Bx\|_2 = 1$. Since Bx' , Bx have the same L_2 norm and x' has a smaller L_2 norm, the supremum must be achieved on H_l . Consider $x = A_{n-l}z$, $z \in S^{n-l-1}$, then

$$\langle x, Y_k^* \rangle^2 = \langle A_{n-l}z, P_l X_k^* \rangle^2 = \langle P_l A_{n-l}z, X_k^* \rangle^2 = \left\langle \sum_{k=l+1}^n z_k X_k, X_k^* \right\rangle^2 = z_k^2.$$

Thus, we have

$$\|B\|^2 = \sup_{z \in S^{n-l-1}} \frac{1}{\|A_{n-l}z\|_2^2} = s_{n+1-l}(A_{n-l})^{-2}.$$

By Theorem 2.6, we have

$$\mathbb{P} \left(\|B\| > \alpha \frac{\sqrt{n}}{l} \right) = \mathbb{P} \left(s_{n+1-l}(A_{n-l}) < \alpha^{-1} \frac{l}{\sqrt{n}} \right) \leq (C\alpha^{-1})^l + \exp(-Cn).$$

Step 1.2.3. Concentration of $\|BX\|$.

By Lemma 2.3, B is independent to X . So we may condition on B such that $\|B\|_{\text{HS}} > \alpha^{-1} \sqrt{\frac{n}{l}}$ and $\|B\| < \alpha \frac{\sqrt{n}}{l}$. By equation (3) and Theorem 2.10,

$$(9) \quad \mathcal{L}(X_i, \beta) \leq cp\beta, \text{ for all } i \in [n].$$

So, we may apply Theorem 2.9 with $\varepsilon = \frac{1}{2}$ and have

$$\begin{aligned}
 (10) \quad \mathbb{P} \left(\|BX\|_2 \leq \beta \alpha^{-1} \sqrt{\frac{n}{l}} \right) &\leq \mathbb{P} (\|BX\|_2 \leq \beta \|B\|_{\text{HS}}) \\
 &\leq (C\beta)^{cr(B)} \leq (C\beta)^{\frac{l}{2\alpha^4}}.
 \end{aligned}$$

Conclusion of step 1. Consider the events,

$$\begin{aligned}
 (11) \quad \mathcal{E}_1 &:= \left\{ A : \|P_l^\perp A_l\| > \alpha \sqrt{l} \right\} \\
 \mathcal{E}_2 &:= \left\{ A : \|B\|_{\text{HS}} < \alpha^{-1} \sqrt{\frac{n}{l}}, \text{ and } \|B\| > \alpha \frac{\sqrt{n}}{l} \right\}.
 \end{aligned}$$

We have shown

$$(12) \quad \begin{aligned} \mathbb{P}(\mathcal{E}_1) &\leq 2 \exp(-C\alpha^2 l) \\ \mathbb{P}(\mathcal{E}_2) &\leq \exp(-C\alpha^2 l) + (C\alpha^{-1})^l + \exp(-Cn) \end{aligned}$$

By conditioning on \mathcal{E}_2^c for all $y \in S^{\ell-1}$ and a vector U defined in (4), we have

$$\mathbb{P} \left(\|A^{-1}U\|_2 < \beta\alpha^{-1} \sqrt{\frac{n}{l}} \middle| \mathcal{E}_2^c \right) < (C\beta)^{\frac{l}{2\alpha^4}}.$$

Step 2: Preparation for the union bound argument.

Now let E_1 (or in fact $\mathbb{R}^{l'}$) be an $l' := \lfloor \kappa l \rfloor$ dimensional coordinate subspace that is spanned by $e_1, \dots, e_{l'}$. We consider an ε -net \mathcal{N}_ε on $S^{l'-1}$ (i.e. S_{E_1}), then $|\mathcal{N}_\varepsilon| \leq (3\varepsilon^{-1})^{l'}$. And for all $y_i \in \mathcal{N}_\varepsilon$, define

$$U_i = U(y_i) := X(y_i) - P_l X(y_i) := A_l y_i - P_l A_l y_i.$$

Step 2.1. $(A_l - P_l A_l)\mathcal{N}_\varepsilon$ is a net on some ellipsoid.

Let

$$E_2 := (A_l - P_l A_l)\mathbb{R}^{l'}, \quad S_2 := (A_l - P_l A_l)S^{l'-1}.$$

By step 1.1, with probability $1 - \exp(-C\alpha^2 l)$, $\|P_l^\perp A_l\| \leq \alpha\sqrt{l}$, i.e., $S_2 \subset \alpha\sqrt{l}B_2^{l'}$.

Also, consider any cap on S_2 of radius δ . Then if $P_l^\perp A_l \mathcal{N}_\varepsilon$ is not a $\delta\sqrt{l}$ -net on S_2 , there exists some $\delta\sqrt{l}$ cap that does not intersect $P_l^\perp A_l \mathcal{N}_\varepsilon$. This means that the pre-image of the cap does not intersect \mathcal{N}_ε . However, if $\|P_l^\perp A_l\| \leq \alpha\sqrt{l}$, then the pre-image contains a cap of radius at least $\frac{\delta}{\alpha}$. Thus, for $\varepsilon = \frac{\delta}{\alpha}$, with probability $1 - \exp(-C\alpha^2 l)$, $P_l^\perp A_l \mathcal{N}_\varepsilon$ is a $\delta\sqrt{l}$ -net on $S_2 \subset \alpha\sqrt{l}B_2^{E_2}$. We denote this $\delta\sqrt{l}$ -net by $\mathcal{N}_\delta := P_l^\perp A_l \mathcal{N}_\varepsilon$.

Step 2.2. Reduction of our objective.

Now, we want to show that for some small choice of κ , $\|A^{-1}U\|_2 \gtrsim \sqrt{\frac{n}{l}}$, for all $U \in S_2$ with high probability. If we can prove this, then together with step 2.1., we have $s_{l'}(A^{-1}) \gtrsim \frac{\sqrt{n}}{l}$ with high probability.

On the event that $P_l^\perp A_l \mathcal{N}_\varepsilon$ forms a $\delta\sqrt{l}$ net on S_2 , we have for all $U \in S_2$, there exists some $U_i \in \mathcal{N}_\delta$ such that $\|U - U_i\|_2 \leq \delta\sqrt{l}$, and

$$\|A^{-1}U\|_2 \geq \|A^{-1}U_i\|_2 - \|A^{-1}(U_i - U)\|_2.$$

For the first term, since we have

$$|\mathcal{N}_\delta| = |\mathcal{N}_\varepsilon| \leq (3\varepsilon^{-1})^{l'} = \left(\frac{3\alpha}{\delta}\right)^{l'}$$

we obtain $\|A^{-1}U_i\|_2 \gtrsim \sqrt{\frac{n}{l}}$, for all U_i with high probability by choosing κ small.

To bound $\|A^{-1}(U_i - U)\|_2$ from above for $\|U_i - U\|_2 \lesssim \sqrt{l}$, we only have to prove

$$\|A^{-1}|_{E_2}\| \lesssim \frac{\sqrt{n}}{l}$$

with high probability.

Step 2.3. Upper bound for $\|A^{-1}|_{E_2}\|$.

Notice that

$$\|A^{-1}|_{E_2}\| \leq \|A^{-1}P_l^\perp A_{l'}\| \cdot \|(P_l^\perp A_{l'})^{-1} : E_2 \rightarrow \mathbb{R}^{l'}\| = \frac{\|A^{-1}P_l^\perp A_{l'}\|}{s_{\min}(P_l^\perp A_{l'})}.$$

We only need to prove for κ small enough:

- (1) $s_{\min}(P_l^\perp A_{l'}) \gtrsim \sqrt{l}$ with high probability.
- (2) $\|A^{-1}P_l^\perp A_{l'}\| \lesssim \sqrt{\frac{n}{l}}$ with high probability.

Step 2.3.1. Lower bound of $s_{\min}(P_l^\perp A_{l'})$.

First, by Theorem 2.5,

$$\mathbb{P}\left(\|P_l^\perp A_{l'}\| \geq \alpha\sqrt{l}\right) \leq 2\exp(-C\alpha^2 l).$$

Next, consider a $\frac{1}{2\alpha^2}$ -net \mathcal{N} on $S^{l'-1}$, then $|\mathcal{N}| \leq (6\alpha^2)^{l'}$. And for all $y_i \in \mathcal{N}$, consider $A_{l'}y_i$ as a random vector. We use an elementary inequality $\mathcal{L}(Z, mt) \leq m\mathcal{L}(Z, t)$ which holds for any $m \in \mathbb{N}$. Consider $\beta < \alpha^{-1}$, then by equation (3) and Theorem 2.10,

$$(13) \quad \mathcal{L}((A_{l'}y_i)_j, \alpha^{-1}) \leq cp\beta \left[\frac{\alpha^{-1}}{\beta} + 1 \right] \leq 2cp\alpha^{-1}, \text{ for all } j \in [n].$$

P_l^\perp is decided by A_{n-l} , which is independent with $A_{l'}y_i$. So we may consider P_l^\perp as a fixed matrix and apply Theorem 2.8 to obtain

$$\mathbb{P}\left(\|P_l^\perp A_{l'}y_i\|_2 \leq \alpha^{-1}\sqrt{l}\right) \leq (C\alpha^{-1})^l.$$

Let $y \in S^{l'-1}$ and choose $y_i \in \mathcal{N}$ with $\|y - y'\|_2 < \frac{1}{2\alpha^2}$. Conditioning on A such that $\|P_l^\perp A_{l'}\| \leq \alpha\sqrt{l}$, then we have

$$(14) \quad \begin{aligned} \|P_l^\perp A_{l'}y\|_2 &\geq \|P_l^\perp A_{l'}y_i\|_2 - \|P_l^\perp A_{l'}\|\|y - y_i\|_2 \\ &\geq \alpha^{-1}\sqrt{l} - \frac{1}{2\alpha^2}\alpha\sqrt{l} = \frac{1}{2\alpha}\sqrt{l} \end{aligned}$$

with probability $1 - (C\alpha^{-1})^l$.

Thus, with a standard union bound argument, we have

$$\mathbb{P}\left(s_{\min}(P_l^\perp A_{l'}) \geq \frac{1}{2\alpha}\sqrt{l}\right) \geq 1 - 2\exp(-C\alpha^2 l) - (6\alpha^2)^{l'} (C\alpha^{-1})^l.$$

Step 2.3.2. Upper bound of $\|A^{-1}P_l^\perp A_{l'}\|$.

Recall for $y \in S^{l'-1}$

$$\|A^{-1}P_l^\perp A_{l'} y\|_2^2 = \|y\|_2^2 + \|A^{-1}P_l A_{l'} y\|_2^2 = 1 + \|BA_{l'} y\|_2^2$$

where

$$\|B\|_{\text{HS}}^2 = \sum_{k=l+1}^n \text{dist}(X_k, H_{l,k})^{-2}.$$

Thus, we only need to show $\|BA_{l'}\| \lesssim \sqrt{\frac{n}{l}}$. We will prove this using Theorem 2.5. To apply Theorem 2.5, we employ an argument that is presented in [8, Section 5.4.1. and Section 13.2.]. This argument provides an upper estimate of $\|B\|_{\text{HS}}$.

Recall that the weak L^p norm of a random variable Z is defined as

$$\|Z\|_{p,\infty} = \sup_{t>0} t \cdot (\mathbb{P}\{|Z| > t\})^{1/p}.$$

Although it is not a norm, it is equivalent to a norm if $p > 1$. In particular, the weak triangle inequality holds:

$$\left\| \sum_i Z_i \right\|_{p,\infty} \leq C(p) \sum_i \|Z_i\|_{p,\infty}$$

where $C(p)$ is bounded above by an absolute constant for $p \geq 2$, see [16], Theorem 3.21.

Now by Theorem 2.7, for any $t > 0$,

$$\mathbb{P}\left\{\text{dist}(X_k, H_{l,k}) \leq t\sqrt{l}\right\} \leq (Ct)^l + \exp(-Cn).$$

Define

$$W_k := \min\left(\text{dist}(X_k, H_{l,k})^{-2}, (t_0\sqrt{l})^{-2}\right)$$

where $t_0 = \frac{C_0 l}{n}$ and C_0 is a small constant depending only on K . Then we have

$$\begin{aligned} \|W_k\|_{l/2,\infty} &= \sup_{t>0} t \cdot (\mathbb{P}\{W_k > t\})^{2/l} \\ &= \sup_{t>0} t^{-2} l^{-1} \cdot \left(\mathbb{P}\left\{W_k^{-\frac{1}{2}} < t\sqrt{l}\right\}\right)^{2/l} \\ (15) \quad &= \sup_{t>t_0} t^{-2} l^{-1} \cdot \left(\mathbb{P}\left\{\text{dist}(X_k, H_{l,k}) < t\sqrt{l}\right\}\right)^{2/l} \\ &\leq \frac{C}{l} + \frac{1}{t_0^2 l} \exp(-C \frac{n}{l}) \\ &\leq \frac{C}{l} + \frac{1}{l} \left(\frac{n^2}{C_0^2 l^2} \exp(-C \frac{n}{l})\right) \leq \frac{C}{l}. \end{aligned}$$

this implies

$$\left\| \sum_{k=l+1}^n W_k \right\|_{l/2, \infty} \leq \frac{C(n-l)}{l} \leq \frac{Cn}{l}.$$

Thus, we have

$$\mathbb{P} \left\{ \sum_{k=l+1}^n W_k > t^2 \frac{n}{l} \right\} \leq (Ct^{-1})^l.$$

On the other hand,

$$\begin{aligned} & \mathbb{P}(\text{there exists } k, W_k \neq \text{dist}(X_k, H_{l,k})^{-2}) \\ (16) \quad & \leq \sum_{k=l+1}^n \mathbb{P} \left\{ \text{dist}(X_k, H_{l,k}) \leq t_0 \sqrt{l} \right\} \\ & \leq (n-l) \left(\left(\frac{CC_0 l}{n} \right)^l + \exp(-Cn) \right) \\ & \leq \exp(-Cl). \end{aligned}$$

So we have

$$\begin{aligned} (17) \quad & \mathbb{P} \left\{ \|B\|_{\text{HS}} > t \sqrt{\frac{n}{l}} \right\} \\ & \leq \mathbb{P} \left\{ \sum_{k=l+1}^n W_k > t^2 \frac{n}{l} \right\} + \mathbb{P}(\text{there exists } k, W_k \neq \text{dist}(X_k, H_{l,k})^{-2}) \\ & \leq (Ct^{-1})^l + \exp(-Cl). \end{aligned}$$

Now, denote

$$\mathcal{E}'_2 := \mathcal{E}_2 \cup \left\{ A : \|B\|_{\text{HS}} > \alpha \sqrt{\frac{n}{l}} \right\}.$$

Applying Theorem 2.5 with $D = B, G = A_{l'}$, $Cs = \frac{1}{2}\alpha$, $Ct = \frac{1}{2}\alpha\sqrt{\frac{l}{l'}}$, we have for α large enough,

$$\begin{aligned} (18) \quad & \mathbb{P} \left\{ \|BA_{l'}\| > \alpha^2 \sqrt{\frac{n}{l}} \middle| (\mathcal{E}'_2)^c \right\} \\ & \leq \mathbb{P} \left\{ \|BA_{l'}\| > \frac{1}{2}\alpha \|B\|_{\text{HS}} + \left(\frac{1}{2}\alpha \sqrt{\frac{l}{l'}} \right) \sqrt{l'} \|B\| \middle| (\mathcal{E}'_2)^c \right\} \\ & \leq 2 \exp(-C\alpha^2 l (\alpha^{-4} + 1)) \leq 2 \exp(-C\alpha^2 l). \end{aligned}$$

So we have

$$\mathbb{P} \left(\|A^{-1} P_l^\perp A_{l'}\| \geq \alpha^2 \sqrt{\frac{n}{l}} \middle| (\mathcal{E}'_2)^c \right) \leq 2 \exp(-C\alpha^2 l).$$

Conclusion of Step 2.

Denote

(19)

$$\begin{aligned}\mathcal{E}_3 &:= \left\{ A : \|A^{-1}|_{E_2}\| \geq 2\alpha^3 \frac{\sqrt{n}}{l} \right\} \cup \mathcal{E}_1 \cup \mathcal{E}'_2, \\ \mathcal{E}_4 &:= \left\{ A : \text{there exists } y_i \in \mathcal{N}_\varepsilon \text{ such that } \|A^{-1}U(y_i)\|_2 \leq \beta\alpha^{-1} \sqrt{\frac{n}{l}} \right\}.\end{aligned}$$

Then we have

$$\mathbb{P}(\mathcal{E}_3) \leq (6\alpha^2)^{l'} (C\alpha^{-1})^l + 4\exp(-C\alpha^2 l) + \exp(-Cl) + \mathbb{P}(\mathcal{E}'_2) + \mathbb{P}(\mathcal{E}_1).$$

Since $|N_\varepsilon| \leq \left(\frac{\alpha\sqrt{l}}{\delta}\right)^{l'-1}$ as we discussed in step 2.1,

$$\mathbb{P}(\mathcal{E}_4 \mid \mathcal{E}_2^c) \leq \left(\frac{3\alpha\sqrt{l}}{\delta}\right)^{l'} (C\beta)^{\frac{l}{2\alpha^4}}$$

Step 3. The union bound argument.

Denote

(20)

$$\begin{aligned}\mathcal{E} := & \left\{ A : \text{there exists } y \in S^{l'-1}, \text{ such that } \|A^{-1}U(y)\|_2 \leq \frac{\beta}{2\alpha} \sqrt{\frac{n}{l}}, \right. \\ & \left. \text{or } \|U(y)\|_2 \geq \alpha\sqrt{l} \right\}.\end{aligned}$$

Choose $\delta = 2\alpha^4\delta = \frac{\beta}{2}$. Let $y \in S^{l'-1}$ and choose $y_i \in \mathcal{N}_\varepsilon$ with $\|y - y'\|_2 < \delta$. If $A \notin \mathcal{E}_3 \cup \mathcal{E}_4$, then $\|U(y)\| \leq \alpha\sqrt{l}$ and

$$\begin{aligned}\|A^{-1}U(y)\|_2 &\geq \|A^{-1}U(y)\|_2 - \|A^{-1}(U(y_i) - U(y))\|_2 \\ (21) \quad &\geq \beta\alpha^{-1} \sqrt{\frac{n}{l}} - 2\alpha^3 \frac{\sqrt{n}}{l} \cdot \delta \\ &\geq \frac{\beta}{2\alpha} \sqrt{\frac{n}{l}}.\end{aligned}$$

Thus, we have $\mathcal{E} \subset \mathcal{E}_3 \cup \mathcal{E}_4$. On the other hand,

$$\begin{aligned}(22) \quad A \in \mathcal{E}^c &\Rightarrow s_{l'}(A^{-1}) \geq \frac{\beta}{2\alpha^2} \frac{\sqrt{n}}{l} \Rightarrow s_{n+1-l'}(A) \leq \frac{2\beta}{\alpha^2} \frac{l}{\sqrt{n}} \\ &\Rightarrow s_{n+1-l'}(A) \leq \frac{4\beta}{\kappa\alpha^2} \frac{l'}{\sqrt{n}}, \text{ for all } l' < \frac{\kappa\tilde{c}n}{2}.\end{aligned}$$

The $\frac{\tilde{c}n}{2}$ in $l' < \frac{\tilde{c}n}{2}$ comes from the requirement that $l \leq \frac{\tilde{c}n}{2}$.

Now let $\beta = \exp(-\alpha^5)$, then $\delta = \frac{1}{4}\alpha^{-4}\exp(-\alpha^5)$. Choose α to be a big enough constant, then

$$\begin{aligned}
 (23) \quad & \mathbb{P}(\mathcal{E}) \\
 & \leq \mathbb{P}(\mathcal{E}_3) + \mathbb{P}(\mathcal{E}_4) \\
 & \leq \left(\frac{3\alpha}{\delta}\right)^{l'} (C\beta)^{\frac{l}{2\alpha^4}} + 7\exp(-C\alpha^2 l) + 2(C\alpha^{-1})^l + \exp(-Cl) + \exp(-Cn) \\
 & \quad + (6\alpha^2)^{l'} (C\alpha^{-1})^l \\
 & \leq ((12\alpha^5 \exp(\alpha^5))^\kappa C \exp(-\frac{1}{2}\alpha))^l + (C\alpha^{2\kappa-1})^l + 9(C\alpha^{-1})^l + \exp(-Cl).
 \end{aligned}$$

Replace l', l by $l, \kappa^{-1}l$, then for a sufficiently small κ depending on α , and $l < \frac{\kappa \tilde{C}n}{2}$, we have

$$\begin{aligned}
 (24) \quad & \mathbb{P}\left(s_{n+1-l}(A) \geq \frac{4\beta}{\kappa\alpha^2} \frac{l}{\sqrt{n}}\right) \\
 & \leq (\exp(-\frac{1}{4}\alpha))^{\kappa^{-1}l} + \left(C\alpha^{-\frac{1}{2}}\right)^{\kappa^{-1}l} + 9(C\alpha^{-1})^{\kappa^{-1}l} + \exp(-Cl) \\
 & \leq (C\alpha^{-\frac{1}{2}})^{\kappa^{-1}l} + \exp(-Cl).
 \end{aligned}$$

Choosing a sufficiently large α , we show that there exist C_1, C_2, C_3 depending only on K, p , such that

$$\mathbb{P}\left(s_{n+1-l}(A) \geq C_1 \frac{l}{\sqrt{n}}\right) \leq \exp(-C_2 l)$$

for all $l \leq C_3 n$. For $l > C_3 n$, the above bound follows from the estimate for $s_1(A)$. So we have for all $1 \leq l \leq n$,

$$\mathbb{P}\left(s_{n+1-l}(A) \geq C_1 \frac{l}{\sqrt{n}}\right) \leq \exp(-C_2 l).$$

□

Remark 3.1. As the proof demonstrates, Assumption 1.3 is satisfied with $s = \beta$ which only depends on p and K .

Remark 3.2. The only place we used the non-degeneracy condition is in the application of 2.9. We expect that the same result holds without the concentration function condition. To remove that condition, the application of Theorem 2.9 on BX must be replaced by showing matrix B does not have a good arithmetic structure with high probability (for arithmetic structure of random matrices and its application, see [4, 7, 10, 12, 8]).

4. DEDUCTION OF COROLLARY 1.10 AND 1.11

Both Theorem 1.10 and 1.11 are direct corollaries of Theorem 1.6.

Proof of Theorem 1.10. Construct an $n \times n$ random matrix J such that its first $n - k$ rows are matrix A and the rest are i.i.d. entries with the same distribution as $A_{i,j}$. Then, by Theorem 1.6, for all $t > 0$ and k between l and n , $s_{n+1-l}(J) \leq \frac{C_1 tl}{\sqrt{n}}$, with probability $1 - \exp(-C_2 tl)$, where C_1, C_2 are constants that depend only on K, p . This implies, with the same probability, there exists an l -dimensional subspace E such that $\|Jy\|_2 \leq \frac{C_1 tl}{\sqrt{n}}$ for all $y \in S_E$.

Let $F := \text{span}\{e_1, \dots, e_{n-k}\}$, then $\|Jy\|_2 \leq \frac{C_1 tl}{\sqrt{n}}$, for all $y \in S_{E \cap F}$ with probability $1 - \exp(-C_2 tl)$. This implies

$$\mathbb{P}_J \left\{ s_{n+1-l}(A) \geq \frac{C_1 tl}{\sqrt{n}} \right\} \leq \exp(-C_2 tl)$$

Then we only need to notice that the above event is independent of the last k rows of J ; thus the probability is also with respect to A . \square

To prove Theorem 1.11, in addition to applying Theorem 1.6, we only need Theorem 2.6 to give a lower bound.

Proof of Theorem 1.11. For the lower bound, denote J as the matrix of the first $n - l$ rows of A . Then we have by Theorem 2.6

$$\begin{aligned} (25) \quad \mathbb{P} \left\{ s_{n+1-l}(A) < \frac{C_1 l}{\sqrt{n}} \right\} &\leq \mathbb{P} \left\{ s_{n+1-l}(J) < \frac{C_1 l}{\sqrt{n}} \right\} \\ &\leq (CC_1)^l + \exp(-Cn) \leq \frac{1}{2} \exp(-\frac{1}{2}C_3 l) + \exp(-Cn) \\ &\leq \frac{1}{2} \exp(-C_3 l) \end{aligned}$$

with some small constant C_3 . The upper bound follows directly with a large t in Theorem 1.6. \square

Note that Theorem 1.11 is a generalization of Theorem 1.3 in [17]. Theorem 1.12 can be proved in the same way as Theorem 1.11.

5. ACKNOWLEDGEMENT

I want to thank my advisor Professor Mark Rudelson for many helpful discussions and advice.

A. PROOF OF THEOREM 2.1

Theorem 2.1. (1) Let D be an $n \times n$ invertible matrix with columns $v_k = De_k$, $k = 1, 2, \dots, n$. Define $v_k^* = (D^{-1})^* e_k$. Then $(v_k, v_k^*)_{k=1}^n$ is a complete biorthogonal system in \mathbb{R}^n .

- (2) Let $(v_k)_{k=1}^n$ be a linearly independent system in an n -dimensional Hilbert space H . Then there exist unique vectors $(v_k^*)_{k=1}^n$ such that $(v_k, v_k^*)_{k=1}^n$ is a biorthogonal system in H . This system is complete.
- (3) Let $(v_k, v_k^*)_{k=1}^n$ be a complete biorthogonal system in a Hilbert space H . Then $\|v_k^*\|_2 = 1/\text{dist}(v_k, \text{span}(v_j)_{j \neq k})$ for $k = 1, 2, \dots, n$.

Proof. (1) follows directly from

$$\langle v_j, v_k^* \rangle = \langle De_j, (D^{-1})^* e_k \rangle = \langle D^{-1} De_j, e_k \rangle = \langle e_j, e_k \rangle = \delta_{j,k}.$$

To prove (2), we use the fact that any basis on a finite dimensional vector space has a unique dual basis. Since H is Hilbert space, the dual basis also belongs to H which forms a biorthogonal system together with the original basis. The completeness follows from the dimension argument.

Since $(v_i, v_i^*)_{i=1}^n$ is a complete biorthogonal system on the Hilbert space H , for any $k = 1, 2, \dots, n$, we have $v_k^* \perp \text{span}\{v_i : i \neq k\}$ and $\{v_i : i = 1, \dots, n, i \neq k\} \cup \{v_k^*\}$ form a basis on H . Thus we have the decomposition

$$v_k = \sum_{i \neq k} a_i v_i + \text{dist}(v_k, \text{span}(v_j)_{j \neq k}) \frac{v_k^*}{\|v_k^*\|_2}.$$

Take inner product with v_k^* at both sides and we have

$$1 = \text{dist}(v_k, \text{span}(v_j)_{j \neq k}) \|v_k^*\|_2,$$

which proves (3). □

REFERENCE

- [1] A. Basak and M. Rudelson. Invertibility of sparse non-hermitian matrices. *arXiv:1507.03525*.
- [2] C. Cacciapuoti, A. Maltsev, and B. Schlein. Local Marchenko-Pastur law at the hard edge of sample covariance matrices. *Journal of Mathematical Physics*, 54(43302), 2012.
- [3] G. Livshyts, G. Paouris, and P. Pivovarov. On sharp bounds for marginal densities of product measures. *arXiv:1507.07949*.
- [4] M. Rudelson. Invertibility of random matrices: norm of the inverse. *Annals of Mathematics*, 168(575-600), 2008.
- [5] Hoi H. Nguyen and Van H. Vu. Normal vector of a random hyperplane. *arXiv:1604.04897*.
- [6] E. Rebrova and K. Tikhomirov. Coverings of random ellipsoids, and invertibility of matrices with i.i.d. heavy-tailed entries. *arXiv:1508.06690*.
- [7] M. Rudelson. Lecture notes on non-asymptotic random matrix theory. *notes from the AMS Short Course on Random Matrices*, 2013.

- [8] M. Rudelson and R. Vershynin. No-gaps delocalization for general random matrices. arXiv:1506.04012.
- [9] M. Rudelson and R. Vershynin. The least singular value of a random square matrix is $o(n^{-1/2})$. *Comptes rendus de l'Académie des sciences - Mathématique*, 346(893-896), 2008.
- [10] M. Rudelson and R. Vershynin. The Littlewood-Offord problem and invertibility of random matrices. *Advances in Mathematics*, 218(600-633), 2008.
- [11] M. Rudelson and R. Vershynin. The smallest singular value of a rectangular random matrix. *Communications on Pure and Applied Mathematics*, 62(1707-1739), 2009.
- [12] M. Rudelson and R. Vershynin. Non-asymptotic theory of random matrices: extreme singular values. *Proceedings of the International Congress of Mathematicians*, Volume III(1576-1602), 2010.
- [13] M. Rudelson and R. Vershynin. Delocalization of eigenvectors of random matrices with independent entries. *Duke Math. Journal*, 164(2507-2538), 2013.
- [14] M. Rudelson and R. Vershynin. Hanson-Wright inequality and sub-gaussian concentration. *Electronic Communications in Probability*, 18(1-9), 2013.
- [15] M. Rudelson and R. Vershynin. Small ball probabilities for linear images of high dimensional distributions. *International Mathematics Research Notices*, 2015(9594-9617), 2014.
- [16] E. M. Stein and G. Weiss. *Introduction to Fourier analysis on Euclidean spaces. Princeton Mathematical Series, No. 32.* 1971.
- [17] S. J. Szarek. Spaces with large distance to l_∞^n and random matrices. *American Journal of Mathematics*, 112(899-942), 1990.
- [18] T. Tao and V. Vu. Inverse Littlewood-Offord theorems and the condition number of random discrete matrices. *Annals of Mathematics*, 169(595-632), 2009.
- [19] T. Tao and V. Vu. Random matrices: The distribution of the smallest singular values. *Geometric and Functional Analysis*, 20(260-297), 2010.